

Tomography by Fourier synthesis

P. Maréchal and A. Elasmaï

Institut de Mathématiques de Toulouse

January 30, 2014

Abstract

We consider a particular approach to the regularization of the inverse problem of computerized tomography. This approach is based on notions pertaining to Fourier synthesis. It refines previous contributions, in which the preprocessing of the data was performed according to the Fourier slice theorem. Since real models must account for the geometrical system response and possibly Compton scattering and attenuation, the Fourier slice theorem does not apply, yielding redefinition of the preprocessing. In general, the latter is not explicit, and must be performed numerically. The most natural choice of preprocessing involves the computation of unstable solutions. A proximal strategy is proposed for this step, which allows for accurate computations and preserves global stability of the reconstruction process.

Warning. This paper was submitted to the Journal of Inverse and Ill-Posed Problems on May 10, 2010, accepted there on November 28, 2010, and however never published.

1 Introduction

From the mathematical viewpoint, tomography consists in solving an ill-posed operator equation of the form $Rf_0 = g$, where f_0 is the unknown image, R is a linear operator (a simplified version of which being the Radon transformation) and g is the data provided by the imaging device (*e.g.* a SPECT or PET camera). The data g is usually referred to as the *sinogram*.

In the ideal model, R is the standard Radon operator, given by

$$(Rf)(\boldsymbol{\theta}, r) = \int f(\mathbf{x}) \delta(r - \langle \boldsymbol{\theta}, \mathbf{x} \rangle) d\mathbf{x}. \quad (1)$$

Here, $\boldsymbol{\theta} = (\cos \phi, \sin \phi)$ is an element of the unit circle in \mathbb{R}^2 , r is a real variable and δ denotes the Dirac *delta function*, so that the

above integral is in fact the (one-dimensional) integral of f over the line determined by the equation $r = \langle \boldsymbol{\theta}, \mathbf{x} \rangle$. In this case, the celebrated *Fourier Slice Theorem* allows for considering the constraints on f_0 imposed by the equation $Rf_0 = g$ as constraints on the Fourier transform of f_0 , these constraints appearing on a radial domain of the Fourier plane. This observation led the authors of [8, 9] to regularize the ill-posed operator equation by means of concepts pertaining to *Fourier synthesis* [5].

Recall that *Fourier synthesis* refers to the generic problem of recovering a function f_0 from a partial and approximate knowledge of its Fourier transform. In [5], the analysis of spectral properties of the truncated Fourier operator led the authors to introduce a regularization principle which can be regarded as a reformulation into a well-posed problem of Fourier interpolation. The original problem of recovering the unknown object is replaced by that of recovering a limited resolution version of it, namely, $\phi_\beta * f_0$, where ϕ_β is some convolution kernel (or point spread function). In [1], inspired by well-known results from the approximation of L^p -functions by means of mollifiers, the authors established variational results for this type of regularization. They regarded ϕ_β as a member of the one-parameter family $\{\phi_\beta | \beta \in \mathbb{R}_+^*\}$ defined by

$$\phi_\beta(\mathbf{x}) := \frac{1}{\beta^2} \phi\left(\frac{\mathbf{x}}{\beta}\right),$$

and studied the behavior of the reconstructed object as β goes to zero.

Notice that mollifiers were also considered in [6]. By writing the *mollified* function $f_\beta := \phi_\beta * f$ as

$$f_\beta(\mathbf{x}) := \int \phi_\beta(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) d\mathbf{y} = \langle \varphi_\beta(\mathbf{x}, \cdot), f \rangle,$$

with $\varphi_\beta(\mathbf{x}, \mathbf{y}) := \phi_\beta(\mathbf{x} - \mathbf{y})$, the duality associated to the underlying inner product is used to approximate f_β via

$$f_\beta(\mathbf{x}) \simeq \langle R^* \psi_\beta(\mathbf{x}), f \rangle = \langle \psi_\beta(\mathbf{x}), Rf \rangle,$$

in which $\psi_\beta(\mathbf{x}) \simeq (R^*)^{-1} \varphi_\beta(\mathbf{x}, \cdot)$. This approach is referred to as the method of *approximate inverses*.

Our approach inherits features from both Fourier synthesis and the approximate inverses. In Section 2, we give a precise description of our methodology in a discrete setting. We shall focus on the problem of Computed Tomography. We shall introduce the notion of *pseudo-commutant* of a matrix with respect to another. In section 3, we discuss various computational strategies, especially concerning what we call *preprocessing of the data*, which consists in applying the aforementioned pseudo-commutant. Finally, in Section 4, we prove the numerical feasibility of our approach by means of reconstructions in emission

tomography, in which the evolution towards realistic *projectors* made it necessary to give up the mere application of the Fourier Slice Theorem. Our numerical experiments will clearly indicate improvement in terms of stability and image quality.

2 A reconstruction methodology

As outlined in the introduction, our aim is to reconstruct a smoothed version $\phi_\beta * f_0$ of the original object f_0 . Recall that, if the operator modelling data acquisition is actually the Radon operator R defined in Equation (1), it is easy to generate the data corresponding to $\phi_\beta * f_0$. As a matter of fact, it results immediately from the Fourier Slice Theorem that

$$R(\phi_\beta * f_0) = R\phi_\beta \circledast Rf_0, \quad (2)$$

where \circledast denotes the convolution with respect to r . If g is an approximation of Rf_0 , then $R\phi_\beta \circledast g$ will be an approximation of $R(\phi_\beta * f_0)$. In terms of operators, this amounts to the existence of an operator Φ_β such that $RC_\beta = \Phi_\beta R$, where C_β is the convolution operator: $C_\beta f := \phi_\beta * f$. Now, realistic models describing the data acquisition in emission or transmission tomography involve operators which are not the exact Radon transformation, and which do not satisfy, in general, Equation (2).

In a recent paper [2], an extension of the regularization by mollification was proposed, in order to cope with operators R for which it is not possible to explicitly find an operator Φ_β such that $RC_\beta = \Phi_\beta R$. The idea consists in defining Φ_β as an operator minimizing

$$X \mapsto \|RC_\beta - XR\|,$$

where $\|\cdot\|$ is some operator norm. In [2], the authors focus on the infinite dimensional setting, while considering general operators. Here, we concentrate on the discrete case. Therefore, from now on, R , C_β and X are matrices of respective sizes $m \times n$, $n \times n$ and $m \times m$, and we shall deal with the minimization of the Frobenius norm the matrix $RC_\beta - XR$ (with respect to X).

Recall that the Frobenius norm of a matrix $M \in \mathbb{R}^{m \times n}$, denoted by $\|M\|_F$, is the Euclidean norm of M regarded as a vector in \mathbb{R}^{mn} , and that the corresponding inner product, denoted by $\langle \cdot, \cdot \rangle_F$, satisfies:

$$\forall M, N \in \mathbb{R}^{m \times n}, \quad \langle M, N \rangle_F = \text{tr}(MN^\top) = \text{tr}(N^\top M).$$

The following theorem can be found in various sources and various forms. We provide a proof for the sake of completeness. Recall that a matrix $X \in \mathbb{R}^{n \times m}$ is the pseudo-inverse M^\dagger of M if and only if it satisfies $MXM = M$, $XM X = X$, $(MX)^\top = MX$ and $(XM)^\top = XM$.

Theorem 2.1 Let A and R be real matrices of size $m \times n$. Then the matrix AR^\dagger minimizes the function $f(X) := \|A - XR\|_F^2 / 2$ over $\mathbb{R}^{m \times m}$. Among all minimizers, it is the one with minimum Frobenius norm.

PROOF. Clearly, f is convex, indefinitely differentiable, and its gradient at X is equal to $(XR - A)R^\top$. Therefore,

$$\nabla f(AR^\dagger) = A(R^\dagger R R^\top - R^\top) = A((R R^\dagger R)^\top - R^\top) = 0.$$

This proves that AR^\dagger minimizes f . Furthermore, since the Frobenius norm is strictly convex, minimizers of f can only differ by matrices in the kernel of the linear mapping $\mathcal{R}: X \mapsto XR$ (from $\mathbb{R}^{m \times m}$ to $\mathbb{R}^{m \times n}$). Now, for all $K \in \mathbb{R}^{m \times m}$,

$$\langle AR^\dagger, K \rangle_F = \text{tr}(AR^\dagger R R^\dagger K^\top) = \text{tr}(AR^\dagger (R^\dagger)^\top (K R)^\top).$$

This implies that every matrix K in the kernel of \mathcal{R} is orthogonal (for $\langle \cdot, \cdot \rangle_F$) to AR^\dagger . The desired conclusion follows by Pythagoras' theorem. ■

Now, letting A be the matrix RC_β , we see that the minimum Frobenius norm minimizer of the matrix function $X \mapsto \|RC_\beta - XR\|_F$ is the matrix

$$\Phi_\beta := RC_\beta R^\dagger \in \mathbb{R}^{m \times m}.$$

In order to be as consistent as possible with our aim, that is, with the reconstruction of the convolution of the original image by our point spread function ϕ_β , we must replace the original sinogram with its transformation by $RC_\beta R^\dagger$.

Let \mathbf{f} and \mathbf{g} denote the discrete versions of f and g , respectively. Consider the decomposition of the generic image \mathbf{f} as the sum of its low and high frequency components:

$$\mathbf{f} = C_\beta \mathbf{f} + (I - C_\beta) \mathbf{f}.$$

We define the reconstructed image as the solution to the following optimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{ll} \text{Minimize} & \frac{1}{2} \|RC_\beta R^\dagger \mathbf{g} - R\mathbf{f}\|^2 + \frac{\alpha}{2} \|H_\beta \mathbf{f}\|^2 \\ \text{subject to} & \mathbf{f} \geq \mathbf{0}, \end{array} \right.$$

in which α is a positive weight and $H_\beta := I - C_\beta$. From the computational viewpoint, the difficult part is the estimation of the *regularized data* $RC_\beta R^\dagger \mathbf{g}$. The reason is, of course, the ill-posedness of the model matrix R , which yields ill-posedness of the computation of $R^\dagger \mathbf{g}$. This difficulty will be addressed in the next section.

3 Computational aspects

In this section, we address the computation of the regularized data, that is, of $RC_\beta R^\dagger \mathbf{g}$. The resolution of problems such as (\mathcal{P}) is quite standard, and will not be discussed here.

Due to the dimension of the involved matrices, it seems unreasonable to actually compute either $RC_\beta R^\dagger$ or R^\dagger . The most natural strategy consists in computing $R^\dagger \mathbf{g}$ as the (unstable) minimum norm least square solution of the original system, and then in applying RC_β to the obtained solution.

It may seem inappropriate to initialize a regularization scheme by computing an unstable solution. It is important, however, to realize that ill-posedness encompasses two aspects (which are related): the numerical inaccuracy induced by the poor conditioning of the system, and the propagation of measurement errors $\delta \mathbf{g}$. We conjecture, at this point, that the former can be dealt with by means of a *proximal* strategy, while the latter is not crucial since the unstable solution $R^\dagger \mathbf{g}$ is post-processed by the smoothing operator C_β (and ultimately by R). In essence, the proximal point algorithm is a fixed point method. The reason for choosing an iterative scheme which, incidentally, has the reputation of being slow, is that it can give very accurate solutions. In fact, it may be used for refining solutions provided by some other method.

In the next paragraph, we review a few aspects of the Proximal Point Algorithm (PPA), applied to the computation of minimum norm least square solutions.

The Proximal Point Algorithm was introduced by Martinet [10] in 1970, in the context of variational inequalities. It was then generalized by Rockafellar [12] to the computation of zeros of maximal monotone operators (a particular case of which being the minimization of a convex function). In our context, that of computing the minimum norm least-square solution of the linear system $R\mathbf{f} = \mathbf{g}$, the PPA consists in the following steps:

1. INITIALIZATION: put $k = 0$ and choose an initial point $\mathbf{f}_0 \in \mathbb{R}^n$.
2. ITERATION: form the sequence $(\mathbf{f}_k)_{k \geq 0}$ according to

$$\mathbf{f}_{k+1} := \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{g} - R\mathbf{f}\|^2 + \frac{1}{2\lambda_k} \|\mathbf{f} - \mathbf{f}_k\|^2 \right\},$$

where $(\lambda_k)_{k \geq 0}$ is a sequence of positive real numbers.

Observe that the function to be minimized in the iteration is strictly convex, smooth and coercive, and differs from the Tikhonov functional only by the subtraction of \mathbf{f}_k in the regularizing part. It has the numerical stability inherent to Tikhonov regularization, and may be computed

by using any routine from either quadratic optimization (*e.g.* the conjugate gradients) or linear algebra (on the corresponding *regularized* normal equation).

In [12], Rockafellar obtained convergence results for the proximal point algorithm, in the general setting of the computation of zeros of maximal monotone operators in real Hilbert spaces. For the sake of simplicity, we restrict attention to the minimization of any lower semi-continuous convex function on \mathbb{R}^n . The proof of the following theorem is deferred to the appendix.

Theorem 3.1 Let $F: \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a lower semi-continuous convex function which is not identically equal to infinity and bounded below, and let \mathbf{x}_0 be any point in \mathbb{R}^n . Consider the sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ defined by

$$\mathbf{x}_{k+1} = \operatorname{argmin} \left\{ F(\mathbf{x}) + \frac{1}{2\lambda_k} \|\mathbf{x} - \mathbf{x}_k\|^2 \mid \mathbf{x} \in \mathbb{R}^n \right\},$$

where $(\lambda_k)_{k \in \mathbb{N}}$ is a sequence of positive numbers. If the series $\sum \lambda_k$ is divergent, then

$$F(\mathbf{x}_k) \rightarrow \eta := \inf \{F(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\} \quad \text{as } k \rightarrow \infty.$$

If in addition the set of minimizers $S := \{\mathbf{x} \in \mathbb{R}^n \mid F(\mathbf{x}) = \eta\}$ is nonempty, then the sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converges to a point $\bar{\mathbf{x}}$ in S .

Now, let F be the function $\mathbf{f} \mapsto \|\mathbf{g} - R\mathbf{f}\|^2/2$. In this case, the set S of minimizers is the affine manifold

$$S = \{\mathbf{f} \in \mathbb{R}^n \mid R^\top R\mathbf{f} = R^\top \mathbf{g}\} = \{R^\dagger \mathbf{g}\} + \ker R.$$

Furthermore, writing the usual first order necessary optimality condition for \mathbf{f}_{k+1} yields

$$\mathbf{f}_{k+1} - \mathbf{f}_k = -\lambda_k R^\top (R\mathbf{f}_{k+1} - \mathbf{g}) \in \operatorname{ran} R^\top = (\ker R)^\perp.$$

This shows that, if $\mathbf{f}_0 = \mathbf{0}$, the limit point of the sequence $(\mathbf{f}_k)_{k \in \mathbb{N}}$ is nothing but $R^\dagger \mathbf{g}$ itself.

Notice that, in the case where R is injective, the set of minimizers of $\|R\mathbf{f} - \mathbf{g}\|$ is reduced to the singleton $\{R^\dagger \mathbf{g}\} = \{(R^\top R)^{-1} R^\top \mathbf{g}\}$. In this case, numerical errors in the proximal iteration are unimportant, since each new iterate can be regarded as a new initial point. However, if R is not injective, a component in the kernel of R may grow due to finite precision in the iteration. In order to cope with this difficulty, the following strategy may be adopted. In practice, the smoothing properties of RC_β are expected to give rise to reasonable perturbation transmission: a small perturbation $\delta \mathbf{f} \in \mathbb{R}^n$ will yield a reasonably small perturbation $RC_\beta \delta \mathbf{f} \in \mathbb{R}^m$. Therefore, replacing R^\dagger with the

usual Tikhonov approximation $(R^\top R + \varepsilon I)^{-1} R^\top$, where I denotes the identity matrix of appropriate dimension and ε is a very small positive parameter, will give a reasonable approximation of the regularized data:

$$RC_\beta(R^\top R + \varepsilon I)^{-1} R^\top \mathbf{g} \approx RC_\beta R^\dagger \mathbf{g}.$$

In the above proximal scheme, the iterations should then be replaced by

$$\mathbf{f}_{k+1} := \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{g} - R\mathbf{f}\|^2 + \frac{\varepsilon}{2} \|\mathbf{f}\|^2 + \frac{1}{2\lambda_k} \|\mathbf{f} - \mathbf{f}_k\|^2 \right\},$$

which will clearly yield convergence to $(R^\top R + \varepsilon I)^{-1} R^\top \mathbf{g}$.

We stress that mathematical convergence is guaranteed whenever the series $\sum \lambda_k$ is divergent, and that this happens in particular if $\lambda_k \equiv \lambda$ with $\lambda > 0$. Notice that the smaller λ_k , the smaller the size of the step $\mathbf{f}_{k+1} - \mathbf{f}_k$. Consequently, it may be appropriate to start the algorithm with large values of λ_k yielding large (but poorly conditioned) steps, and then let λ_k decrease in order to have smaller but well-conditioned steps. Moreover, as indicated in [7], numerical accuracy in early iterations may be irrelevant: what really matters is the limit of the proximal sequence. We terminate this section with a few comments.

(1) If the size of R is not prohibitively large, the computation of R^\dagger may be performed by means of the *singular value decomposition* (SVD). The problems of tomographic reconstruction encountered in practice make it difficult or even impossible to perform the SVD. Approximate pseudo-inverses may then be obtained by truncating the SVD. Such approximate solution may be used as initial points for the PPA.

(2) As observed in [7], the proximal iteration with constant sequence $\lambda_k \equiv \lambda$ belongs to the class of *fixed point methods*, along with the algorithms of Jacobi, Gauss-Seidel, SOR and SSOR. It is easy to check that $R^\dagger \mathbf{g}$ satisfies the fixed point equation $\mathbf{f} = h(\mathbf{f})$, with

$$h(\mathbf{f}) := B\mathbf{f} + \mathbf{c}, \quad B := (I + \lambda R^\top R)^{-1} \quad \text{and} \quad \mathbf{c} := \lambda(I + \lambda R^\top R)^{-1} R^\top \mathbf{g}.$$

Clearly, h is a contraction and, if R is injective, then $R^\top R$ is positive definite and h is a strict contraction.

4 Numerical results

We now illustrate our reconstruction approach by a few numerical simulation. To begin with, we use a 64×64 image of the Shepp-Logan phantom, as shown in Figure 1. The sinogram is obtained by means of an operator R which accounts for the geometrical system response,

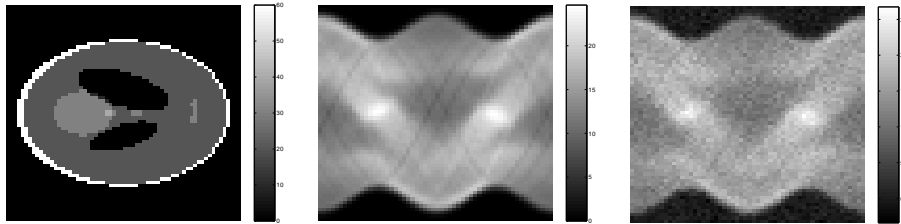


Figure 1: The shepp-Logan phantom (left), the corresponding sinogram (middle) and the noisy sinogram (right).

as described in [3, 11] (the resolution in the projection depends on the distance to the detector). The simulated acquisition was performed with 64 bins and 64 angular values evenly spaced over 360 degrees. An independent Poisson noise was added to the sinogram, with parameter equal to the pixel value. (The total number of counts in the noisy sinogram is equal to 50065.)

The reconstructions shown in Figure 2 were obtained by solving Problem (\mathcal{P}) . The data is preprocessed according to the strategy described in the previous section. For comparison, reconstructions without preprocessing of the data are shown.

In practice, the positivity constraint can often be neglected: removing it from problem (\mathcal{P}) turns out to give images which are essentially positive. The advantage of this is that we deal with purely quadratic optimization, which allows for stability analysis as well as the use of the *backward error* stopping criterion, as in [9]: the reconstructions are performed using the conjugate gradients algorithm, with construction of the Galerkin tridiagonal matrix.

It is clear that preprocessing enhances the quality of the reconstruction. These simulations corroborate the anticipated relevance of our approach, and attests its numerical feasibility. It is interesting to note that, as conjectured in Section 3, the unstable step consisting in computing $R^\dagger \mathbf{g}$ does not damage the whole process: the application of C_β to $R^\dagger \mathbf{g}$ is sufficient to ensure stability of the reconstruction.

Table 3 displays the *normalized quadratic error*

$$\mathcal{E}(\mathbf{f}) := \frac{\|C_\beta \mathbf{f}_0 - \mathbf{f}\|}{\|C_\beta \mathbf{f}_0\|}$$

for various solutions \mathbf{f} : in the first line, \mathbf{f} is reconstructed without preprocessing of the data, while in the second line, \mathbf{f} is obtained without preprocessing. In the last line, the value (given for reference) is that corresponding to the reconstruction by Filtered Back-Projection (FBP). Here, \mathbf{f}_0 denotes the original object.

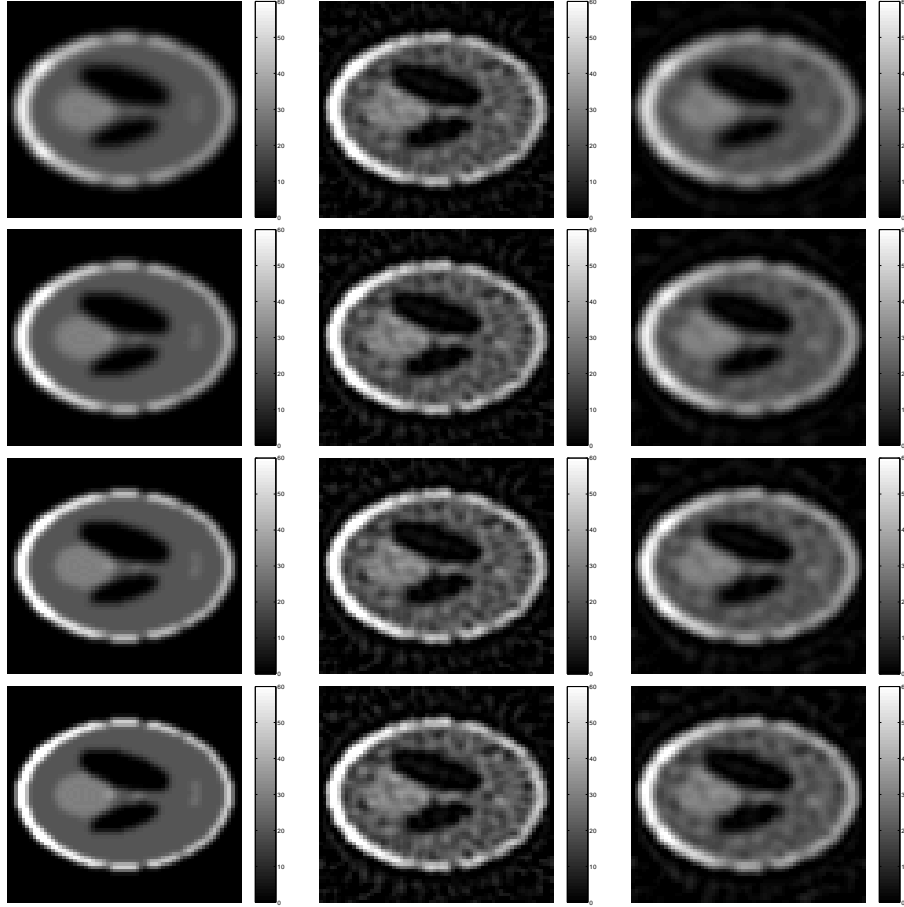


Figure 2: Images in the left column are mere convolutions of the original phantom image by the objective point-spread function (here a Hann filter is used), with cutoff frequency equal, in Nyquist units, to 0.5 (first line), 0.6 (second line), 0.7 (third line) and 0.8 (last line). The second and third columns show the corresponding reconstructions, respectively without and with preprocessing of the data.

	Cutoff frequency			
	0.5	0.6	0.7	0.8
without preprocessing	0.347786	0.301052	0.287429	0.291660
with preprocessing	0.118515	0.140563	0.168467	0.202067
with FBP	0.868203	0.879254	0.887703	0.891690

Figure 3: Normalized quadratic error $\mathcal{E}(\mathbf{f})$ for various cutoff frequencies and three reconstruction methods: Fourier synthesis without and with preprocessing (respectively first and second lines), and the Filtered Back-Projection (last line).

5 Conclusion

We have considered an extension of the regularization by mollification of the problem of computerized tomography. This was motivated by the fact that realistic models differ from the standard Radon transform: the latter cannot be used if the system response accounts for geometrical aspects as well as Compton scattering and attenuation.

The data corresponding to the objective of our reconstruction process (namely a smoothed version of the original object) can be computed numerically by means of the proximal point algorithm, which operates even when the dimension of the system matrix makes it difficult or impossible to use the SVD.

Simulations have shown both the numerical feasibility of our regularization scheme and its efficiency, in terms of robustness and image quality. The unstable computation of the pseudo-inverse of the data fits in a regularization strategy without affecting its global stability.

Such a strategy may be used to apply the regularization by mollification to many other imaging techniques. More generally, the proximal algorithm may provide the user with an efficient tool for preprocessing the data in accordance with the objective of the reconstruction, whenever the latter objective is a linear transform of the original image.

6 Appendix: proof of Theorem 3.1

We start with a finite dimensional version of Opial's lemma:

Lemma 6.1 *Let $(\mathbf{x}_k)_{k \in \mathbb{N}}$ be an \mathbb{R}^n -valued sequence, and let S be a nonempty subset of \mathbb{R}^n . Suppose that*

- (i) *for all $\mathbf{x} \in S$, the sequence $(\|\mathbf{x}_k - \mathbf{x}\|)_{k \in \mathbb{N}}$ has a limit;*
- (ii) *every cluster point of (\mathbf{x}_k) belongs to S .*

Then (\mathbf{x}_k) converges to a point $\bar{\mathbf{x}}$ in S .

PROOF. Condition (i) implies that (\mathbf{x}_k) is bounded, which implies in turn that the sequence has at least one cluster point, say $\bar{\mathbf{x}}$. We shall prove that every cluster point $\tilde{\mathbf{x}}$ must coincide with $\bar{\mathbf{x}}$. Let $(\mathbf{x}_{k_j})_{j \in \mathbb{N}}$ be a subsequence converging to $\bar{\mathbf{x}}$ and $(\mathbf{x}_{k'_j})_{j \in \mathbb{N}}$ be a subsequence converging to $\tilde{\mathbf{x}}$. Condition (ii) implies that both $\bar{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ belong to S . Condition (i) again implies that

$$\|\mathbf{x}_k - \bar{\mathbf{x}}\|^2 - \|\mathbf{x}_k - \tilde{\mathbf{x}}\|^2$$

has a limit. Developing the above expression then shows that the sequence $\langle \mathbf{x}_k, \bar{\mathbf{x}} - \tilde{\mathbf{x}} \rangle$ has a limit, from which we deduce that

$$\lim_{j \rightarrow \infty} \langle \mathbf{x}_{k_j}, \bar{\mathbf{x}} - \tilde{\mathbf{x}} \rangle = \lim_{j \rightarrow \infty} \langle \mathbf{x}_{k'_j}, \bar{\mathbf{x}} - \tilde{\mathbf{x}} \rangle.$$

Therefore, $\langle \bar{\mathbf{x}}, \bar{\mathbf{x}} - \tilde{\mathbf{x}} \rangle = \langle \tilde{\mathbf{x}}, \bar{\mathbf{x}} - \tilde{\mathbf{x}} \rangle$, which immediately yields $\bar{\mathbf{x}} = \tilde{\mathbf{x}}$. ■

Recall that the *effective domain* of a convex function F is defined to be the set

$$\text{dom } F := \{\mathbf{x} \in \mathbb{R}^n \mid F(\mathbf{x}) < \infty\},$$

and that the *subdifferential* of F at a point \mathbf{x} is the (closed convex) set

$$\partial F(\mathbf{x}) := \{\boldsymbol{\xi} \in \mathbb{R}^n \mid \forall \mathbf{y} \in \mathbb{R}^n, F(\mathbf{y}) \geq F(\mathbf{x}) + \langle \boldsymbol{\xi}, \mathbf{y} - \mathbf{x} \rangle\}.$$

Members $\boldsymbol{\xi}$ of the subdifferential are called *subgradients*, and the inequality in the definition of $\partial F(\mathbf{x})$ is referred to as the *subgradient inequality*.

PROOF OF THEOREM 3.1. For all $\mathbf{x} \in \mathbb{R}^n$,

$$F(\mathbf{x}_{k+1}) + \frac{1}{2\lambda_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \leq F(\mathbf{x}) + \frac{1}{2\lambda_k} \|\mathbf{x} - \mathbf{x}_k\|^2.$$

Taking $\mathbf{x} = \mathbf{x}_k$ yields

$$F(\mathbf{x}_{k+1}) + \frac{1}{2\lambda_k} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \leq F(\mathbf{x}_k).$$

This shows that the sequence $F(\mathbf{x}_0), F(\mathbf{x}_1), \dots$ decreases, and since F is bounded below, the sequence converges to some real limit $l \geq \eta := \inf \{F(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$. In order to obtain the first assertion of the theorem, we must prove that actually $l = \eta$. We start by proving the following inequality:

$$\forall \mathbf{x} \in \text{dom } F, \quad \|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \geq 2\lambda_k (F(\mathbf{x}_{k+1}) - F(\mathbf{x})). \quad (3)$$

The optimality of \mathbf{x}_{k+1} reads:

$$\mathbf{0} \in \partial F(\mathbf{x}_{k+1}) + \frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\lambda_k},$$

that is, $\mathbf{x}_k - \mathbf{x}_{k+1} \in \lambda_k \partial F(\mathbf{x}_{k+1})$. In other words, there exists $\boldsymbol{\xi}$ in $\partial F(\mathbf{x}_{k+1})$ such that $\mathbf{x}_k - \mathbf{x}_{k+1} = \lambda_k \boldsymbol{\xi}$. Now, for all $\mathbf{x} \in \text{dom } F$, we have:

$$\begin{aligned}
& \|\mathbf{x}_k - \mathbf{x}\|^2 \\
&= \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 + \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 + 2 \langle \mathbf{x}_k - \mathbf{x}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x} \rangle \\
&\geq \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 + 2 \langle \mathbf{x}_k - \mathbf{x}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x} \rangle \\
&= \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 + 2\lambda_k \langle \boldsymbol{\xi}, \mathbf{x}_{k+1} - \mathbf{x} \rangle \\
&\geq \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 + 2\lambda_k (F(\mathbf{x}_{k+1}) - F(\mathbf{x})),
\end{aligned}$$

in which the last inequality stems from the subgradient inequality. Thus (3) is clear. Next, since $F(\mathbf{x}_{k+1}) \geq l$ for all k , (3) implies that

$$\forall \mathbf{x} \in \text{dom } F, \quad \|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \geq 2\lambda_k (l - F(\mathbf{x})).$$

Summing the above for $k = 0, \dots, n$ shows that, for all $\mathbf{x} \in \text{dom } F$,

$$\|\mathbf{x}_0 - \mathbf{x}\|^2 - \|\mathbf{x}_{n+1} - \mathbf{x}\|^2 \geq 2 \left(\sum_{k=0}^n \lambda_k \right) (l - F(\mathbf{x})).$$

Since the series $\sum \lambda_k$ is divergent, we must have $l \leq F(\mathbf{x})$ for all $\mathbf{x} \in \text{dom } F$, which eventually show that $l \leq \eta$.

It remains to prove that, assuming $S := \text{argmin } F \neq \emptyset$, the proximal sequence (\mathbf{x}_k) converges to a point $\bar{\mathbf{x}} \in S$. Since $F(\mathbf{x}_{k+1}) - F(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in S$, (3) shows that

$$\forall \mathbf{x} \in S, \quad \|\mathbf{x}_k - \mathbf{x}\|^2 \geq \|\mathbf{x}_{k+1} - \mathbf{x}\|^2.$$

Thus $\|\mathbf{x}_k - \mathbf{x}\|$ decreases as k increases, which obviously implies that $\|\mathbf{x}_k - \mathbf{x}\|$ goes to a limit. It follows that the sequence (\mathbf{x}_k) is bounded. Now, by lower semi-continuity, every cluster point $\bar{\mathbf{x}}$ of (\mathbf{x}_k) (there is at least one since the sequence is bounded) satisfies

$$F(\bar{\mathbf{x}}) \leq \lim_{j \rightarrow \infty} F(\mathbf{x}_{k_j}) = \lim_{k \rightarrow \infty} F(\mathbf{x}_k) = l.$$

Thus every cluster point of (\mathbf{x}_k) belongs to S , and Opial's lemma shows that the sequence (\mathbf{x}_k) actually converges to a point $\bar{\mathbf{x}}$ in S . ■

References

- [1] N. ALIBAUD, P. MARÉCHAL and Y. SAESOR, *A variational approach to the inversion of truncated Fourier operators*, Inverse Problems **25**, 2009.

- [2] X. BONNEFOND and P. MARÉCHAL, *A variational approach to the inversion of some compact operators*, Pacific Journal of Optimization **5**(1), pp. 97-110.
- [3] A.R. FORMICONI, A. PUPI, A. PASSERI, *Compensation of spatial system response in SPECT with conjugate gradient reconstruction technique*, Phys. Med. Biol. **34**, pp. 69-84, 1989.
- [4] P. M. KOULIBALY, I. BUVAT, M. PELEGRINI and G. EL FAKHRI, *Correction de la perte de résolution avec la profondeur*, Revue de l'ACOMEN, **4**, pp. 114-120, 1998.
- [5] A. LANNES, S. ROQUES and M.-J. CASANOVE, *Stabilized reconstruction in signal and image processing; Part I: partial deconvolution and spectral extrapolation with limited field*, J. Mod. Opt. **34**, pp. 161-226, 1987.
- [6] A.K. LOUIS and P. MAASS, *A mollifier method for linear operator equations of the first kind*, Inverse Problems **6**, pp. 427-440, 1990.
- [7] P. MARÉCHAL, A. RONDEPIERRE, *A proximal approach to the inversion of ill-conditioned matrices*, C. R. Acad. Sci. Paris, Ser. I **347**, pp. 1435-1438, 2009.
- [8] P. MARÉCHAL, D. TOGANE and A. CELLER, *A new reconstruction methodology for computerized tomography: FRECT (Fourier Regularized Computed Tomography)*, IEEE, Trans. Nucl. Sc., **47**, pp. 1595-1601, 2000.
- [9] D. MARIANO-GOULART, P. MARÉCHAL, L. GIRAUD, S. GRATTON and M. FOURCADE, *A priori selection of the regularization parameters in emission tomography by Fourier synthesis*, Computerized Medical Imaging and Graphics, **31**, pp. 502-509, 2007.
- [10] B. MARTINET, *Régularisation d'inéquations variationnelles par approximations successives*, Revue Française d'Informatique et de Recherche Opérationnelle, pp. 154-159, 1970.
- [11] A. PASSERI, A.R. FORMICONI and U. MELDOLESI, *Physical modelling (geometrical system response, Compton scattering and attenuation) in brain SPECT using the conjugate gradient reconstruction method*, Phys. Med. Biol., **37**, pp. 1727-1744, 1992.
- [12] R.T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, **14** (5), pp.877-898, 1976.